Audio Engineering Society

Convention Paper

Presented at the 118th Convention
2005 May 28–31 Barcelona, Spain

Pseudostereophony Revisited

Christof Faller[1]

[1] *Audiovisual Communications Laboratory, EPFL Lausanne, Switzerland*

Correspondence should be addressed to Christof Faller (`christof.faller@epfl.ch`)

ABSTRACT

Conventional stereophonic processes allow playback of mono audio signals with a stereo effect. The stereo effect is limited to mimicking ambience or signal independent left/right separation and thus no realistic sound stage is reproduced. This paper proposes two techniques for converting old mono recordings to two or more channel stereo signals with a realistic sound stage and ambience. One technique is fully automatic and imposes the auditory spatial image of a given modern stereo recording onto a corresponding old mono recording. The other technique is based on manual input of a sound engineer to generate a desired sound stage and ambience. The underlying mono-to-stereo synthesis process is the same as has been recently proposed for use in low bitrate audio coding.

1. INTRODUCTION

There is a long history in techniques attempting to "enhance" mono signals to create a spatial impression, i.e. to generate a signal pair or more channels evoking some kind of spatial impression. Such techniques are often called "pseudostereophonic" processes. In [1] a scheme was proposed where a lowpass filtered version of the mono signal is given to one loudspeaker and a highpass filtered version to the other loudspeaker. Another technique uses complementary comb filters for generating left and right signals [2]. The use of allpass filters instead of comb filters, resulting in a stereo effect with less coloration artifacts, has been proposed in [3]. The use of a reverberation chamber with one loudspeaker emitting the mono signal and two microphones generating left and right signals was described in [4] and [5]. Another scheme gives the mono signal to both loudspeakers and adds an attenuated and delayed version of the mono signal to one loudspeaker and the same phase inverted attenuated and delayed signal to the other loudspeaker [2, 6]. In [7] the use of time-variant controllable filters, controlled by properties of the mono signal, was proposed. A more thorough review on these pseudosterephonic processes is given in [8]. In all the described techniques, the spatial distribution of the auditory events is independent of where the sound was originally picked up. Even more limiting, it is not possible to explicitly control the location of different instruments in the recording. The stereo effect is arbitrary and spatialization can not be associated with the different instruments.

Another class of techniques has been proposed to convert two channel stereo signals into multi-channel surround signals. One way of doing this is to apply a conventional matrixing decoder to (non-matrixed) stereo signals. For example, Dolby Prologic [9] decoders can be used for playing back stereo signals as 5.1 surround [10] signals. Another technique has been proposed [11], which operates in the short-time Fourier transform (STFT) domain. This technique detects ambient signal portions in time and frequency and plays part of these back over the 5.1 rear channels. However, these techniques have not the same aim as the techniques proposed in this paper since they require a given stereo signal.

The techniques proposed in this paper have the aim, similar as pseudostereophonic processes, to generate two or more channel stereo signals given mono recordings. The fundamental difference between the conventional techniques and the techniques described in this paper is that the enhanced mono signals evoke a specific auditory spatial image as desired. For example, old orchestral recordings are processed such that the generated stereo signal evokes a realistic auditory spatial image in a listener, i.e. the different instruments of the orchestra are reproduced as auditory events at realistic defined locations.

Different recordings of the same classical music score usually feature the same instruments. One of the proposed techniques, denoted *automatic pseudostereophonic process*, takes advantage of this and imposes spatial signal cues obtained from a modern stereo recording to the old mono recording. The result is a stereo version of the old recording with a similar auditory spatial image as the modern stereo recording.

The second proposed technique, denoted *manual pseudostereophonic process*, requires input from a recording engineer during the process of generation of the stereo signal. We developed a graphical user interface (GUI) based software for this purpose as is described later.

The proposed techniques are based on a stereo synthesis process which generates a stereo signal by means of synthesizing inter-channel differences between the channels. The automatic pseudostereophonic process obtains the inter-channel differences from a given stereo recording. For the manual pseudostereophonic process the inter-channel differences are generated by means of using a GUI-based software with which the desired auditory spatial image is determined.

The paper is organized as follows. Section 2 describes the process which generates stereo signals given mono signals. Both proposed techniques utilize this mono-to-stereo process. The automatic pseudostereophonic process in described in Section 3 and the manual pseudostereophonic process is described in Section 4. Section 5 describes the experiments we carried out and informal subjective impressions. The conclusions are presented in Section 6.

2. MONO-TO-STEREO SYNTHESIS

Recently, parametric coding of stereo signals [12, 13, 14, 15, 16, 17, 18, 19], using a similar stereo synthesis process as used here, was proposed. These schemes can achieve good audio quality, indicating that also the proposed pseudostereophonic processes can achieve good audio quality.

These coding techniques are based on a mono-to-stereo synthesis process. Inter-channel cues are synthesized as a function of time in different subbands. The bandwidths of the subbands are chosen according to the spectral resolution of the human auditory system.

Summing localization [8] implies that perceptually relevant audio channel differences for a loudspeaker signal channel pair are the *inter-channel time difference* (ICTD) and *inter-channel level difference* (ICLD). ICTD and ICLD can be related to the perceived direction of auditory events [8, 20, 21]. Other auditory spatial image attributes, such as apparent source width [22] and listener envelopment [23], can be related to *interaural coherence* (IC) [24, 22]. For loudspeaker pairs in the front or back of a listener, the interaural coherence is often directly related to the *inter-channel coherence* (ICC) [25] which is thus considered as third audio channel difference measure by the mono-to-stereo synthesis process.

The following measures are used for ICTD, ICLD,

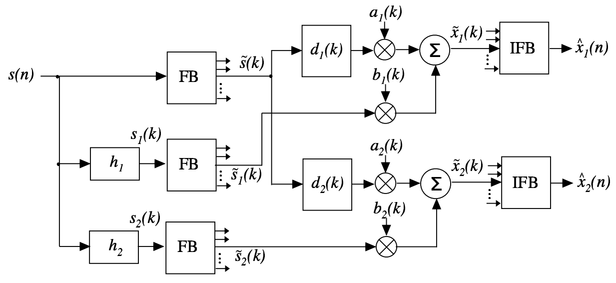

Fig. 1: Mono-to-stereo synthesis: The given mono signal and two filtered versions thereof are decomposed into a number of subbands. The subbands are processed such that specific time differences, level differences, and coherence cues appear between the two output channels.

and ICC for corresponding subband signals $\tilde{x}_1(k)$ and $\tilde{x}_2(k)$ of two audio channels:

- ICTD [samples]:

$$\tau_{12}(k) = \arg\max_d \{\Phi_{12}(d,k)\}\,, \tag{1}$$

with a short-time estimate of the normalized cross-correlation function

$$\Phi_{12}(d,k) = \frac{p_{\tilde{x}_1\tilde{x}_2}(d,k)}{\sqrt{p_{\tilde{x}_1}(k-d_1)p_{\tilde{x}_2}(k-d_2)}}\,, \tag{2}$$

where

$$\begin{aligned} d_1 &= \max\{-d, 0\} \\ d_2 &= \max\{d, 0\}\,, \end{aligned} \tag{3}$$

and $p_{\tilde{x}_1\tilde{x}_2}(d,k)$ is a short-time estimate of the mean of $\tilde{x}_1(k-d_1)\tilde{x}_2(k-d_2)$.

- ICLD [dB]:

$$\Delta L_{12}(k) = 10\log_{10}\left(\frac{p_{\tilde{x}_2}(k)}{p_{\tilde{x}_1}(k)}\right)\,. \tag{4}$$

- ICC:

$$c_{12}(k) = \max_d |\Phi_{12}(d,k)|\,. \tag{5}$$

Note that the absolute value of the normalized cross-correlation is considered and $c_{12}(k)$ has a range of $[0,\ 1]$.

The mono-to-stereo synthesis process is shown in Figure 1 [26]. The input signal $s(n)$ is filtered with two filters modeling late reverberation. The resulting three signals, $s(n)$, $s_1(n)$, and $s_2(n)$, are converted to a subband domain. The delays and scale factors for modifying the subband signals (Figure 1) are computed as [26]:

$$\begin{aligned} a_1 &= \sqrt{\frac{1-A+B}{C}} \\ a_2 &= \sqrt{\frac{A-1+B}{C}} \\ b_1 &= \sqrt{\frac{(A+1-B)p_{\tilde{s}}(k)}{Cp_{\tilde{s}_1}(k)}} \\ b_2 &= \sqrt{\frac{(A+1-B)p_{\tilde{s}}(k)}{Cp_{\tilde{s}_2}(k)}} \\ d_1 &= \max\{-\tau_{12}, 0\} \\ d_2 &= \max\{\tau 12, 0\}\,, \end{aligned} \tag{6}$$

with

$$\begin{aligned} A &= 10^{\frac{\Delta L_{12}(k)}{10}} \\ B &= \sqrt{(1-A)^2 + 4Ac_{12}^2(k)} \\ C &= 2(1+A)\,. \end{aligned} \tag{7}$$

Note that these parameters are computed such that a pair of subbands has the specific inter-channel cues ICTD, ICLD, and ICC.

Both proposed pseudostereophonic processes use different means of obtaining ICTD, ICLD, and ICC which are then applied to generate a stereo signal given the old mono recording.

3. AUTOMATIC PSEUDOSTEREOPHONIC PROCESS

The automatic pseudostereophonic process generates the spatial cues as used by the mono-to-stereo synthesis automatically, given a stereo recording of the same music that is to be converted to stereo. Figure 2 illustrates how the spatial cues are generated. The spatial cues of the given stereo recording are estimated. Furthermore, the delay between the mono and stereo recording is estimated adaptively in time. The estimated spatial cues are delayed such that they are in sync relative to the mono recording. Finally, the stereo signal is generated by using the

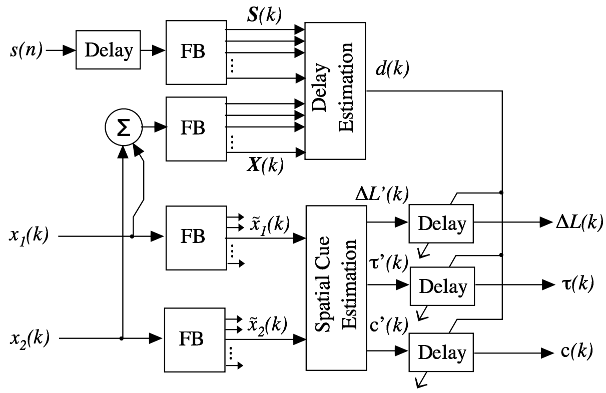

Fig. 2: Scheme for estimation and time alignment of the spatial cues. The cues are estimated from a different stereo recording and delayed according to an estimate of the delay between the old mono and corresponding other stereo recording.

obtained spatial cues for converting the mono signal to stereo (as described in Section 2).

The delay between the mono signal, $s(n)$, and the left/right sum of the stereo signal, $x(n) = x_1(n) + x_2(n)$, is estimated as follows. $s(n)$ and $x(n)$ are converted to the subband domain. The short-time estimate of the power of all the subbands is computed. The vectors with the subband power estimates are denoted $\mathbf{S}(k)$ and $\mathbf{X}(k)$, where k is the subband time index. Temporal and spectral similarity is measured by computing the normalized vector cross-correlation function,

$$\gamma(d) = \frac{\mathrm{E}\{\mathbf{S}(k) \cdot \mathbf{X}(k-d)\}}{\sqrt{\mathrm{E}\{\mathbf{S}(k) \cdot \mathbf{S}(k)\}\mathrm{E}\{\mathbf{X}(k-d) \cdot \mathbf{X}(k-d)\}}}\,, \tag{8}$$

where $\mathrm{E}\{.\}$ denotes mathematical expectation, $\cdot$ is the vector-dot-product operator, and d is the time lag index. Since the delay between $s(n)$ and $x(n)$ is likely to vary in time, a short-time estimate of (8) is computed by

$$\gamma(k,d) = \frac{a_{12}(k,d)}{\sqrt{a_{11}(k,d)a_{22}(k,d)}}\,, \tag{9}$$

where

$$\begin{aligned} a_{12}(k,d) &= \alpha\mathbf{S}(k) \cdot \mathbf{X}(k-d) \\ &\quad +(1-\alpha)a_{12}(k-1,d)\,, \\ a_{11}(k,d) &= \alpha\mathbf{S}(k-d) \cdot \mathbf{S}(k-d) \\ &\quad +(1-\alpha)a_{11}(k-1,d)\,, \\ a_{22}(k,d) &= \alpha\mathbf{X}(k) \cdot \mathbf{X}(k) \\ &\quad +(1-\alpha)a_{22}(k-1,d)\,, \end{aligned}$$

and $\alpha \in [0,1]$ determines the time-constant of the exponentially decaying estimation window

$$T = \frac{1}{\alpha f_s}\,, \tag{10}$$

where f_s denotes the (downsampled) subband sampling frequency. The delay is estimated as the lag of the maximum of the normalized cross-correlation function,

$$\tau(k) = \arg\max_d \gamma(k,d)\,. \tag{11}$$

Note that the time resolution of the computed delay is limited by the subband sampling interval $1/f_s$. The top two panels of Figure 3 show the short-time power spectra $\mathbf{S}(k)$ and $\mathbf{X}(k)$ as a function of time. The bottom panel of Figure 3 shows $\gamma(k,d)$ and $\tau(k)$. The simulation was carried out with a classical mono recording from the 1930's and a corresponding newer stereo recording. Note that the old mono recording (top panel of Figure 3) has less audio bandwidth than the stereo recording (middle panel of Figure 3).

The normalization of the cross-correlation function is introduced in order to get an estimate of the similarity (coherence), defined as the maximum value of the instantaneous normalized cross-correlation function,

$$c_{sx}(k) = \max_d \gamma(k,d)\,. \tag{12}$$

At this point not used, but $c_{sx}(k)$ not close to one indicates the the old mono recording and stereo recording are not very similar spectrally. In this case, one could apply different cues than the estimated cues for minimizing artifacts.

Noise and rumbling in the old recording Often old mono recordings include a substantial amount of noise and rumbling. When the described automatic pseudostereophonic process is applied to such recordings the quality is lower than expected. That is, because noise will be "randomly spatialized" to positions where the non-noise signal components

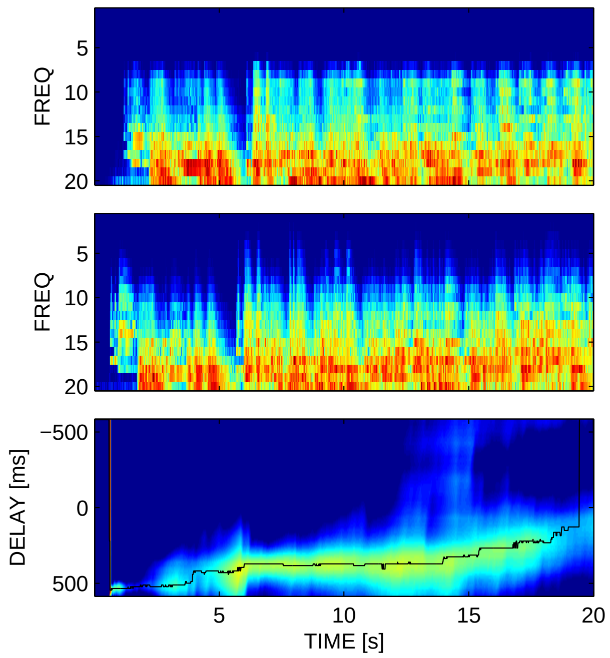

Fig. 3: Power spectrum of mono signal $\mathbf{S}(k)$ (top) and of the left/right sum of the stereo signal $\mathbf{X}(k)$ (middle). The normalized vector cross-correlation $\gamma(k, d)$ (bottom gray) and the estimated delay $\tau(k)$ (bottom black).

of the stereo recording were spatialized. Thus, noise will not be statically centered, but its spatialization changes randomly as a function of time. This results in perceived instability of the auditory spatial image.

To reduce this negative effect, we estimate noise in the mono recording and suppress it prior to application of the mono-to-stereo synthesis. The proposed technique not only estimates static noise, but also "clicks", "pops", or "rumbling noise". The noise is estimated by comparing the power spectrum (subband power values) of the mono recording to the power spectrum of the mono downmix of the stereo recording. Additionally, at frequencies where noise is suppressed different spatial cues are applied.

The algorithm comprises the following steps:

1. The signals $s(n)$ and the downmixed signal $x(n) = x_1(n) + x_2(n)$ are scaled such that their power is the same. (Possibly time-adaptive scaling).

2. The subband power vectors $\mathbf{S}(k)$ and $\mathbf{X}(k)$ are computed.

3. The noise subband power values are estimated by
$$\mathbf{V}(k) = \max\{\mathbf{S}(k) - \mathbf{X}(k), \mathbf{0}\}\,, \qquad (13)$$
where the max operation is applied to each vector element individually and $\mathbf{0}$ is a vector with zero-elements. The rationale behind (13) is to declare signal components which are stronger in the mono recording than in the stereo recording as noise. These are also the signal components for which the spatial cues of the stereo recording may be wrong with respect to the mono recording.

4. A vector with subband scale factors is computed:
$$\mathbf{W}(k) = \sqrt{\frac{\mathbf{X}(k)}{\mathbf{X}(k) + \mathbf{V}(k)}}\,, \qquad (14)$$
where the division and square-root are carried out individually for each vector element. Before the inverse filterbank in Fig. 1 the subbands are scaled with the corresponding scale factors in $\mathbf{W}(k)$. Note that this type of signal modification (frequency dependent scaling) is often used in noise suppression and speech enhancement algorithms [27, 28, 29].

5. For subbands with a substantial amount of noise (e.g. scale factor of more than -10 dB) the ICTD and ICLD are set to zero and the ICC is set to a relatively small value (e.g. 0.5).

4. MANUAL PSEUDOSTEREOPHONIC PROCESS

Figure 4 shows the graphical user interface of the software that is used for the manual pseudostereophonic process. The human operator selects different regions within the time-frequency representation of the given mono recording. Each of the regions is assigned an index of the auditory event that it corresponds to. For each of the auditory events the direction and width is defined. For aiding correct identification of auditory event signal components the software can play back each region separately, such that the operator can modify the regions until they represent the desired signal components.

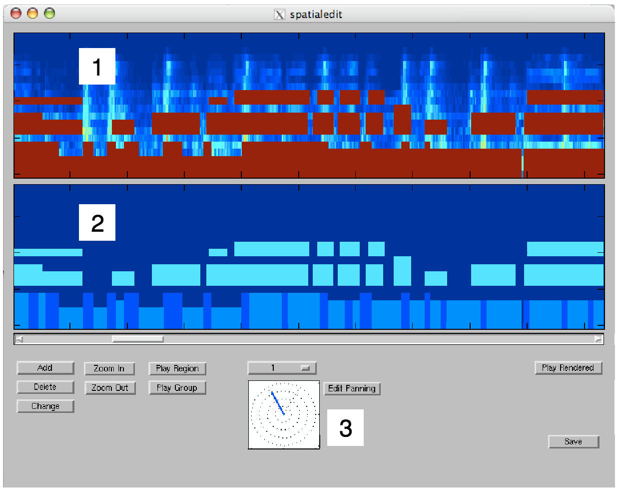

Fig. 4: Software for the manual pseudostereophonic process: 1: The short-time power spectrum is shown, 2: Regions are selected and assigned to different auditory events, 3: For each auditory event the direction and width is defined.

Once all the regions are defined, i.e. the signal components belonging to different auditory events are assigned corresponding indices, the stereo signal is generated with the desired rendering parameters. All non-selected signal portions are treated as ambient sound. For each auditory event and for the ambient sound the ICTD, ICLD, and ICC cues are chosen by the operator such that the resulting stereo signal evokes in a listener the desired auditory spatial image.

In the following, we are describing how the ICLD and ICC are determined for each auditory event and the ambient sound. At this point we do not use ICTD for the manual pseudostereophonic process and thus ICTD is not discussed here. The perceived direction of an auditory event appearing when amplitude panning is applied follows approximately the stereophonic law of sines [30],

$$\frac{\sin\phi}{\sin\phi_0} = \frac{a_1 - a_2}{a_1 + a_2}\,, \tag{15}$$

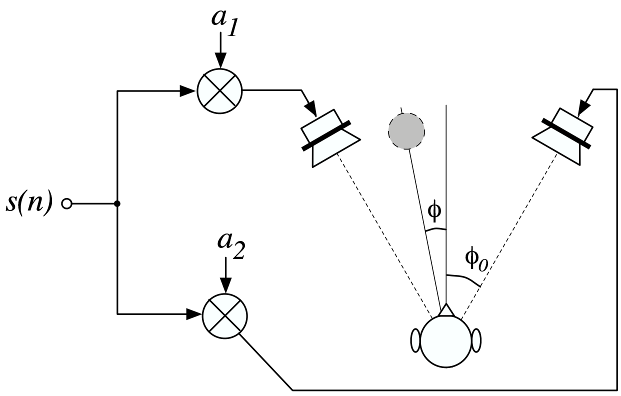

Fig. 5: Definitions of scale factors and angles for the stereophonic law of sines (15).

where $0° \le \phi_0 \le 90°$ is the angle between the forward axis and the two loudspeakers, ϕ is the corresponding angle of the auditory event, and a_1 and a_2 are scale factors determining ICLD. The ICLD as a function of ϕ is

$$\mathrm{ICLD} = 20\log_{10}\frac{a_2}{a_1} = 20\log_{10}\frac{\sin\phi_0 - \sin\phi}{\sin\phi_0 + \sin\phi}\,. \tag{16}$$

Angles and scale factors are illustrated in Figure 5. The relation between ICLD and ϕ is shown in Figure 6 for a standard stereo listening setup with $\phi_0 = 30°$.

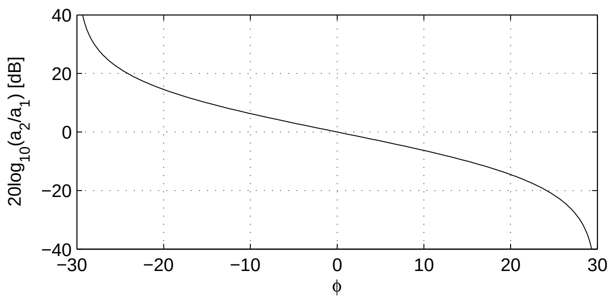

Fig. 6: The relation between auditory event angle ϕ and ICLD, i.e. $20\log_{10}(a_2/a_1)$, for the stereophonic law of sines.

In [31] a panning law considering an improved head model compared to the stereophonic law of sines is derived. The result was a "stereophonic law of tangents" which is similar to another earlier proposed law [32] but for different listening conditions. Amplitude panning and auditory event direction perception is discussed in more detail in [33]. We are using the stereophonic law of sines (16) to determine ICLD.

To control the width of auditory events, the left and right signals are de-correlated after the amplitude panning law has been applied. The less similar the left and right signals are, the more wide is the corresponding auditory event. As a measure for similarity ICC is used.

The desired angle and width determines the ICLD and ICC with which the signal portion of each auditory event index is to be rendered. For the ambient sound, we usually choose ICLD $= 0$ and a relatively small ICC.

5. SUBJECTIVE IMPRESSIONS AND DISCUSSION

5.1. Automatic pseudostereophonic process

In addition to conducting the first performance of Mahler's Ninth with the Vienna Philharmonic, Bruno Walter also made the first recording of the score. This is considered a remarkable phonographic document of the last century. (Gustav Mahler, Bruno Walter, Angel Records, Audio CD June 6, 1989).

As a corresponding stereo recording we took another well respected recording of Mahler's Ninth, a performance of Claudio Abbado with the Berliner Philharmoniker. (Gustav Mahler, Claudio Abbado, Deutsche Grammophon, Audio CD June 11, 2002).

The result of the the automatic pseudostereophonic process is impressive. The music is spatialized and the directions of the instruments in the processed mono recording approximately match the corresponding directions of instruments in the stereo recording. Also ambience is successfully mimicked.

Nevertheless, the automatic pseudostereophonic process not always performs as good as desired. It has difficulty with very transient music (e.g. piano), probably due to the relatively low time resolution that is used. For improving this problem it may be necessary to isolate transients in the stereo recording and perfectly align the corresponding side information with the old recording. Another issue is noise in the old recording and special consideration of the noise by suppressing it or spatializing it differently may further improve the scheme.

5.2. Manual pseudostereophonic process

We applied the manual pseudostereophonic process to the Beatles song "Love Me Do". (Beatles 1, Capitol Records, Audio CD November 14, 2000). This song has relatively few instruments facilitating identification of different instruments by means of the GUI-based software. The result is a stereo version of Love Me Do with specific virtual source positions and amount of ambience determined by the operator of the software.

The manual pseudostereophonic process is easy to apply and yields satisfactory results in cases when the used time frequency representation is able to effectively separate the relevant signal parts. At this point we use just relatively few (e.g. 20) subbands with a frequency resolution motivated by perception. However, to further improve the proposed technique it would be desirable to include functionality into the software to isolate and spatialize harmonic sounds. Also precise isolating of transients is desirable.

6. CONCLUSIONS

We proposed a new class of pseudostereophonic processes. While previous approaches are limited to adding a feeling of ambience or space to the mono signal, the proposed techniques aim at spatializing the mono signal such that a specific auditory spatial image is perceived. That is, not only ambience is added but also the sources in the mono signal are rendered to specific directions. One way of achieving this, is to use a stereo recording of the same music (e.g. classical stereo recording) and impose the auditory spatial image information from the stereo recording onto the mono recording. The other way of achieving this, is to use a GUI-based software to identify signal components which are to be rendered to specific directions.

ACKNOWLEDGMENTS

Thanks to Thibaut Ajdler and David Pardes for suggesting famous old classical mono recordings with corresponding newer stereo recordings.

7. REFERENCES

[1] W. H. Janovsky, "An apparatus for three-dimensional reproduction in electroacoustical presentations," *German Federal Republic Patent No. 973570*, 1948, (cited in Blauert 1997).

[2] H. Lauridsen, "Nogle forsog med forskellige former rumakustic gengivelske," *Ingenioren*, vol. 47, pp. 906, 1954, (cited in Schroeder 1961 and Blauert 1997).

[3] M. R. Schroeder, "Improved quasi-stereophony and "colorless" artificial reverberation," *J. Acoust. Soc. Am.*, vol. 33, pp. 1061–1064, 1961.

[4] M. R. Schroeder, "An artificial stereophonic effect obtained from a single audio signal," *J. Acoust. Soc. Am.*, vol. 6, pp. 74–79, 1958.

[5] J. P. A. Lochner and W. de V. Keet, "Stereophonic and quasi-stereophonic reproduction," *J. Acoust. Soc. Am.*, vol. 32, pp. 392–401, 1960.

[6] H. Lauridsen and F. Schlegel, "Stereofonie und richtungsdiffuse Klangwiedergabe [Stereophony and directionally diffuse reproduction of sound]," *Gravesaner Blätter*, vol. 5, pp. 28–50, 1956, (cited in Blauert 1997).

[7] F. Enkl, "Die übertragung räumlicher Schallfeldstrukturen über einen kanal mit hilfe unterschwelliger pilotfrequenzen [The transmission of spatial sound field structures over one channel aided by pilot tones below the threshold]," *Elektron. Rdsch.*, vol. 12, pp. 347–349, 1958.

[8] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge, Massachusetts, USA, revised edition, 1997.

[9] R. Dressler, "Dolby Surround Prologic II Decoder - Principles of operation," Tech. Rep., Dolby Laboratories, 2000, www.dolby.com/tech/.

[10] Rec. ITU-R BS.775, *Multi-Channel Stereophonic Sound System with or without Accompanying Picture*, ITU, 1993, http://www.itu.org.

[11] C. Avendano and J.-M. Jot, "Ambience extraction and synthesis from stereo signals for multichannel audio up-mix," in *Proc. ICASSP, Orlando, Florida*, May 2002, vol. 2, pp. 1957–1960.

[12] C. Faller and F. Baumgarte, "Binaural Cue Coding: A novel and efficient representation of spatial audio," in *Proc. ICASSP*, May 2002, vol. 2, pp. 1841–1844.

[13] C. Faller and F. Baumgarte, "Binaural Cue Coding applied to stereo and multi-channel audio compression," in *Preprint 112th Conv. Aud. Eng. Soc.*, May 2002.

[14] F. Baumgarte and C. Faller, "Binaural Cue Coding - Part I: Psychoacoustic fundamentals and design principles," *IEEE Trans. on Speech and Audio Proc.*, vol. 11, no. 6, Nov. 2003.

[15] C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and applications," *IEEE Trans. on Speech and Audio Proc.*, vol. 11, no. 6, Nov. 2003.

[16] E. Schuijers, W. Oomen, A. C. den Brinker, and A. J. Gerrits, "Advances parametric coding for high-quality audio," in *Proc. MPCA*, Nov. 2002.

[17] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," in *Preprint 114th Conv. Aud. Eng. Soc.*, Mar. 2003.

[18] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegard, "Low complexity parametric stereo coding," in *Preprint 117th Conv. Aud. Eng. Soc.*, May 2004.

[19] J. Engdegard, H. Purnhagen, J. Roden, and L. Liljeryd, "Synthetic ambience in parametric stereo coding," in *Preprint 117th Conv. Aud. Eng. Soc.*, May 2004.

[20] G. Theile and G. Plenge, "Localization of lateral phantom sources," *J. Audio Eng. Soc.*, vol. 25, no. 4, pp. 196–200, 1977.

[21] V. Pulkki, "Localization of amplitude-panned sources II: Two- and three-dimensional panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 753–757, 2001.

[22] T. Okano, L. L. Beranek, and T. Hidaka, "Relations among interaural cross-correlation coefficient ($IACC_E$), lateral fraction (LF_E), and apparent source width (asw) in concernt halls," *J. Acoust. Soc. Am.*, vol. 104, no. 1, pp. 255–265, July 1998.

[23] M. Morimoto and Z. Maekawa, "Auditory spaciousness and envelopment," in *Proc. 13th Int. Congr. on Acoustics*, Belgrade, 1989, vol. 2, pp. 215–218.

[24] J. S. Bradley, "Comparison of concert hall measurements of spatial impression," *J. Acoust. Soc. Am.*, vol. 96, no. 6, pp. 3525–3535, 1994.

[25] K. Kurozumi and K. Ohgushi, "The relationship between the cross-correlation coefficient of two-channel acoustic signals and sound image quality), and apparent source width (asw) in concernt halls," *J. Acoust. Soc. Am.*, vol. 74, no. 6, pp. 1726–1733, Dec. 1983.

[26] C. Faller, "Parametric multi-channel audio coding: Synthesis of coherence cues," *IEEE Trans. on Speech and Audio Proc.*, 2003, (submitted Dec. 2003, accepted).

[27] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE trans. Acoust. Speech Sig. Processing*, vol. 27, no. 2, pp. 113–120, Nov. 1979.

[28] Y. Ephrain and D. Malah, "Speech enhancement using optimal non-linear spectral amplitude estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Sig. Processing (Boston)*, 1983, pp. 1118–1121.

[29] W. Etter and G. S. Moschytz, "Noise reduction by noise-adaptive spectral magnitude expansion," *J. Audio Eng. Soc.*, vol. 42, pp. 341–349, May 1994.

[30] B. B. Bauer, "Phasor analysis of some stereophonic phenomena," *J. Acoust. Soc. Am.*, vol. 33, pp. 1536–1539, Nov. 1961.

[31] J. C. Bennett, K. Barker, and F. O. Edeko, "A new approach to the assessment of stereophonic sound system performance," *J. Audio Eng. Soc.*, vol. 33, no. 5, pp. 314–321, May 1985.

[32] B. Bernfeld, "Attempts for better understanding of the directional stereophonic listening mechanism," in *Preprint 44th Conv. Aud. Eng. Soc.*, Feb. 1973.

[33] V. Pulkki, "Localization of amplitude-panned sources I: Stereophonic panning," *J. Audio Eng. Soc.*, vol. 49, no. 9, pp. 739–752, 2001.